GenCore version 4.5

OM protein - protein search, using sw model

Run on: March 6, 2002, 13:38:49 ; Search time 13.36 Seconds
(without alignments)
1772.865 Million cell updates/sec

Title: US-09-405-504A-25
Perfect score: 3372
Sequence: 1 MRAPGAGAASVVSLALLWLL.....HYLPLNEAVYTRICSGAFAL 646

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3062	90.8	646	1	FATP_RAT	P97849 rattus norv
2	3054	90.6	646	1	FATP_MOUSE	Q60714 mus musculu
3	1046.5	31.0	620	1	VLCS_RAT	P97524 rattus norv
4	1044.5	31.0	620	1	VLCS_HUMAN	O14975 homo sapien
5	1032.5	30.6	620	1	VLCS_MOUSE	O35488 mus musculu
6	699	20.7	623	1	FAT1_YEAST	P38225 saccharomyc
7	390.5	11.6	522	1	CAIC_ECOLI	P31552 escherichia
8	301.5	8.9	2560	1	PPS2_BACSU	P39846 bacillus su
9	298.5	8.9	544	1	4CL1_PETCR	P14912 petroselinu
10	297.5	8.8	544	1	4CL2_PETCR	P14913 petroselinu
11	293.5	8.7	3587	1	SRF1_BACSU	P27206 bacillus su
12	290.5	8.6	547	1	4CL1_TOBAC	O24145 nicotiana t
13	287.5	8.5	3587	1	SRF2_BACSU	Q04747 bacillus su
14	281.5	8.3	561	1	4CL3_ARATH	Q9s777 arabidopsis
15	281	8.3	563	1	4CL1_ORYSA	P17814 oryza sativ
16	280	8.3	548	1	YDID_ECOLI	P38135 escherichia
17	278.5	8.3	545	1	4CL2_SOLTU	P31685 solanum tub
18	277.5	8.2	545	1	4CL1_SOLTU	P31684 solanum tub
19	272.5	8.1	553	1	4CL_VANPL	O24540 vanilla pla
20	268.5	8.0	542	1	4CL2_TOBAC	O24146 nicotiana t
21	268	7.9	561	1	LCFA_ECOLI	P29212 escherichia
22	259.5	7.7	661	1	ACSA_COPCI	O13440 coprinus ci
23	256.5	7.6	537	1	4CL_PINTA	P41636 pinus taeda
24	256.5	7.6	651	1	ACSA_MYCTU	O69635 mycobacteri
25	249	7.4	6486	1	TYCC_BACBR	O30409 b tyrocidin
26	245	7.3	561	1	4CL1_ARATH	Q42524 arabidopsis
27	245	7.3	562	1	4CL2_SOYBN	P31687 glycine max
28	244.5	7.3	4451	1	GRSB_BACBR	P14688 b gramicidi
29	241.5	7.2	3587	1	TYCB_BACBR	O30408 b tyrocidin
30	237.5	7.0	572	1	ACSA_BACSU	P39062 bacillus su
31	237	7.0	562	1	LCFA_HAEIN	P46450 haemophilus
32	235.5	7.0	556	1	4CL2_ARATH	Q9s725 arabidopsis
33	235	7.0	628	1	PRPE_ECOLI	P77495 escherichia
34	234	6.9	628	1	PRPE_SALTY	P55912 salmonella
35	233	6.9	683	1	ACS2_YEAST	P52910 saccharomyc
36	232	6.9	684	1	ACS2_KLULA	Q9y7b5 kluyveromyc
37	232	6.9	805	1	ACVS_STRCL	Q01757 streptomyce
38	230	6.8	670	1	ACSA_EMENI	P16928 emericella
39	230	6.8	5255	1	BACA_BACLI	O68006 b bacitraci
40	228.5	6.8	6359	1	BACC_BACLI	O68008 b bacitraci
41	227	6.7	543	1	FAT2_YEAST	P38137 saccharomyc
42	226.5	6.7	675	1	ACSA_CANAL	O94049 candida alb
43	226.5	6.7	2555	1	PPS3_BACSU	P39847 bacillus su
44	225.5	6.7	2561	1	PPS1_BACSU	P39845 bacillus su
45	224.5	6.7	669	1	ACSA_PENCH	P36333 penicillium

ALIGNMENTS

```
RESULT   1
FATP_RAT
ID   FATP_RAT       STANDARD;      PRT;   646 AA.
AC   P97849;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).
GN   SLC27A1 OR FATP.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98041635; PubMed=9375787;
RA   Schaap F.G., Hamers L., van der Vusse G.J., Glatz J.F.C.;
RT   "Molecular cloning of fatty acid-transport protein cDNA from rat.";
RL   Biochim. Biophys. Acta 1354:29-34(1997).
CC   -!- FUNCTION: INVOLVED IN TRANSLOCATION OF LONG-CHAIN FATTY ACIDS
CC       ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
CC       AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
CC       IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
CC       TRIGLYCERIDE SYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U89529; AAC53424.1; -.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Glycoprotein; Lipid transport; Transmembrane.
FT   TRANSMEM     14     34       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    140    160       POTENTIAL.
FT   TRANSMEM    169    189       POTENTIAL.
FT   TRANSMEM    293    313       POTENTIAL.
FT   TRANSMEM    537    557       POTENTIAL.
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    393    393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    518    518       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   646 AA;  71283 MW;  C450CF174CC2EB87 CRC64;
```

Query Match 90.8%; Score 3062; DB 1; Length 646;
Best Local Similarity 89.5%; Pred. No. 2e-231;
Matches 578; Conservative 30; Mismatches 38; Indels 0; Gaps 0;

```
Qy       1 MRAPGAGAASVVSLALLWLLGLPWTWSAAAALGVYVGSGGWRFLRIVCKTARRDLFGLSV 60
           || |||| ||| || |||||||||||||||| ||||||||||||||||||||||||||||
Db       1 MRTPGAGTASVASLGLLWLLGLPWTWSAAAAFGVYVGSGGWRFLRIVCKTARRDLFGLSV 60

Qy      61 LIRVRLELRRHQRAGHTIPRIFQAVVQRQPERLALVDAGTGECWTFAQLDAYSNAVANLF 120
           |||||||||||:||| ||||||||| |||||||||||| :| |||||||| |||||||||
Db      61 LIRVRLELRRHRRAGDTIPRIFQAVAQRQPERLALVDASSGICWTFAQLDTYSNAVANLF 120

Qy     121 RQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRREPLAFCLGTSGAKALI 180
           |||||||||||:|||||||||||||||||||: |||||||||||||||||| |||||
Db     121 LQLGFAPGDVVAVFLEGRPEFVGLWLGLAKAGVVAALLNVNLRREPLAFCLGTSAAKALI 180

Qy     181 FGGEMVAAVAEVSGHLGKSLIKFCSGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKGM 240
           :|||| |||||||  |||||:|||||||||| :|||| |||||:| || | |||| | |||
Db     181 YGGEMAAAVAEVSEQLGKSLLKFCSGDLGPESVLPDTQLLDPMLAEAPTTPLAQAPGKGM 240

Qy     241 DDRLFYIYTSGTTGLPKAAIVVHSRYYRMAAFGHHAYRMQAADVLYDCLPLYHSAGNIIG 300
           |||||||||||||||||||||||||||||:||||||:| |:| ||||||||||||||||:|
Db     241 DDRLFYIYTSGTTGLPKAAIVVHSRYYRIAAFGHHSYSMRANDVLYDCLPLYHSAGNIMG 300

Qy     301 VGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRL 360
           ||||:|||||||||||||||||||:|||||||||||||||||||:|||: ||||||||
Db     301 VGQCIIYGLTVVLRKKFSASRFWDDCVKYNCTVVQYIGEICRYLLRQPVRDVERRHRVRL 360

Qy     361 AVGNGLRPAIWEEFTERFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHVYPIR 420
           |||||||||||||||: |||||||||||||||||||||||||||||||||| ||||||
Db     361 AVGNGLRPAIWEEFTQGFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILTHVYPIR 420

Qy     421 LVKVNEDTMELLRDAQGLCIPCQAGEPGLLVGQINQQDPLRRFDGYVSESATSKKIAHSV 480
           |||||||||| |||:|||||||| ||||||||||||||||||||||:|||:|||||||
Db     421 LVKVNEDTMEPLRDSQGLCIPCQPGEPGLLVGQINQQDPLRRFDGYVSDSATNKKIAHSV 480

Qy     481 FSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVY 540
           | ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     481 FRKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEAVLSRLLGQTDVAVY 540

Qy     541 GVAVPGVEGKAGMAAVADPHSLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQ 600
           ||||||||||:||||:||||: |||||::|||||||| ||:|||||||||||||||||||
Db     541 GVAVPGVEGKSGMAAIADPHNQLDPNSMYQELQKVLASYAQPIFLRLLPQVDTTGTFKIQ 600

Qy     601 KTRLQREGFDPRQTSDRLFFLDLKQGHYLPLNEAVYTRICSGAFAL 646
           |||||||||||||||||||||||| ||||:| |: |||:| |:|
Db     601 KTRLQREGFDPRQTSDRLFFLDLKQGRYLPLDERVHARICAGDFSL 646
```

```
RESULT   2
FATP_MOUSE
ID   FATP_MOUSE     STANDARD;      PRT;   646 AA.
AC   Q60714;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   LONG-CHAIN FATTY ACID TRANSPORT PROTEIN (FATP).
GN   SLC27A1 OR FATP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS;
RX   MEDLINE=95042740; PubMed=7954810;
RA   Schaffer J.E., Lodish H.F.;
RT   "Expression cloning and characterization of a novel adipocyte long
RT   chain fatty acid transport protein.";
RL   Cell 79:427-436(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98438516; PubMed=9765271;
RA   Hui T.Y., Frohnert B.I., Smith A.J., Schaffer J.E., Bernlohr D.A.;
RT   "Characterization of the murine fatty acid transport protein gene and
RT   its insulin response sequence.";
RL   J. Biol. Chem. 273:27420-27429(1998).
CC   -!- FUNCTION: INVOLVED IN TRANSLOCATION OF LONG-CHAIN FATTY ACIDS
CC       ACROSS THE PLASMA MEMBRANE. MAY PLAY A PIVOTAL ROLE IN REGULATING
CC       AVAILABLE LONG-CHAIN FATTY ACID SUBSTRATES FROM EXOGENOUS SOURCES
CC       IN TISSUES UNDERGOING HIGH LEVELS OF BETA-OXIDATION OR
CC       TRIGLYCERIDE SYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PLASMA MEMBRANE.
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN SKELETAL MUSCLE, HEART
CC       AND FAT. LOWER LEVELS IN BRAIN, KIDNEY, LUNG AND LIVER. NO
CC       EXPRESSION IN SPLEEN OR INTESTINE.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U15976; AAC71060.1; -.
DR   EMBL; AF023258; AAC69640.1; -.
DR   EMBL; AF023256; AAC69640.1; JOINED.
DR   EMBL; AF023257; AAC69640.1; JOINED.
DR   MGD; MGI:1347098; Slc27a1.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Glycoprotein; Lipid transport; Transmembrane.
FT   TRANSMEM     14     34       POTENTIAL.
FT   TRANSMEM    115    135       POTENTIAL.
FT   TRANSMEM    140    160       POTENTIAL.
FT   TRANSMEM    169    189       POTENTIAL.
FT   TRANSMEM    293    313       POTENTIAL.
FT   TRANSMEM    537    557       POTENTIAL.
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    393    393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    518    518       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   646 AA;  71276 MW;  910B92BA8D985B4C CRC64;
```

```
  Query Match             90.6%;  Score 3054;  DB 1;  Length 646;
  Best Local Similarity   89.5%;  Pred. No. 8.6e-231;
  Matches  578;  Conservative   30;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 MRAPGAGAASVVSLALLWLLGLPWTWSAAAALGVYVGSGGWRFLRIVCKTARRDLFGLSV 60
           ||||||| ||| |||||| |||||||||||  |||| ||||||||||||||||||||||
Db       1 MRAPGAGTASVASLALLWFLGLPWTWSAAAAFCVYVGGGGWRFLRIVCKTARRDLFGLSV 60

Qy      61 LIRVRLELRRHQRAGHTIPRIFQAVVQRQPERLALVDAGTGECWTFAQLDAYSNAVANLF 120
           |||||||||||:||| ||| ||||| :|||||||||| :| |||||||| |||||||||
Db      61 LIRVRLELRRHRRAGDTIPCIFQAVARRQPERLALVDASSGICWTFAQLDTYSNAVANLF 120

Qy     121 RQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRREPLAFCLGTSGAKALI 180
           |||||||||||:|||||||||||||||||||: |||||||||||||||||| |||||
Db     121 RQLGFAPGDVVAVFLEGRPEFVGLWLGLAKAGVVAALLNVNLRREPLAFCLGTSAAKALI 180

Qy     181 FGGEMVAAVAEVSGHLGKSLIKFCSGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKGM 240
           :|||| |||||||  |||||:|||||||||| ||||| ||||:| || | |||| | |||
Db     181 YGGEMAAAVAEVSEQLGKSLLKFCSGDLGPESILPDTQLLDPMLAEAPTTPLAQAPGKGM 240

Qy     241 DDRLFYIYTSGTTGLPKAAIVVHSRYYRMAAFGHHAYRMQAADVLYDCLPLYHSAGNIIG 300
           |||||||||||||||||||||||||||||:|||||:| |:|||||||||||||||||:|
Db     241 DDRLFYIYTSGTTGLPKAAIVVHSRYYRIAAFGHHSYSMRAADVLYDCLPLYHSAGNIMG 300

Qy     301 VGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRL 360
           ||||:|||||||||||||||||||:|||||||||||||||||||:|||: |:||||||
Db     301 VGQCVIYGLTVVLRKKFSASRFWDDCVKYNCTVVQYIGEICRYLLRQPVRDVEQRHRVRL 360

Qy     361 AVGNGLRPAIWEEFTERFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHVYPIR 420
           |||||||||||||||:|||| ||||||||||||||||||||||||||||| ||||||
Db     361 AVGNGLRPAIWEEFTQRFGVPQIGEFYGATECNCSIANMDGKVGSCGFNSRILTHVYPIR 420
```

```
Qy    421 LVKVNEDTMELLRDAQGLCIPCQAGEPGLLVGQINQQDPLRRFDGYVSESATSKKIAHSV 480
          |||||||||| |||::|||||||| |||||||||||||||||||||||:|||:|||||||
Db    421 LVKVNEDTMEPLRDSEGLCIPCQPGEPGLLVGQINQQDPLRRFDGYVSDSATNKKIAHSV 480

Qy    481 FSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVY 540
          | ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    481 FRKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEAVLSRLLGQTDVAVY 540

Qy    541 GVAVPGVEGKAGMAAVADPHSLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQ 600
          |||||||||||||||:|||| |||::|||||||| ||||||||||||||||||||||||
Db    541 GVAVPGVEGKAGMAAIADPHSQLDPNSMYQELQKVLASYARPIFLRLLPQVDTTGTFKIQ 600

Qy    601 KTRLQREGFDPRQTSDRLFFLDLKQGHYLPLNEAVYTRICSGAFAL 646
          |||||||||||||||||||||||| |:|||:| |: |||:| |:|
Db    601 KTRLQREGFDPRQTSDRLFFLDLKQGRYVPLDERVHARICAGDFSL 646
```

```
RESULT   3
VLCS_RAT
ID   VLCS_RAT       STANDARD;      PRT;   620 AA.
AC   P97524;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
DE   FATTY-ACID-COA LIGASE).
GN   FACVL1 OR VLACS.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=WISTAR; TISSUE=Liver;
RX   MEDLINE=97094763; PubMed=8939997;
RA   Uchiyama A., Aoyama T., Kamijo K., Uchida Y., Kondo N., Orii T.,
RA   Hashimoto T.;
RT   "Molecular cloning of cDNA encoding rat very long-chain acyl-CoA
RT   synthetase.";
RL   J. Biol. Chem. 271:30360-30365(1996).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D85100; BAA12722.1; -.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM    107    127       POTENTIAL.
FT   TRANSMEM    262    282       POTENTIAL.
SQ   SEQUENCE   620 AA;  70693 MW;  6CF9362DC3805526 CRC64;
```

```
  Query Match             31.0%;  Score 1046.5;  DB 1;  Length 620;
  Best Local Similarity   38.6%;  Pred. No. 4.9e-74;
  Matches  242;  Conservative  100;  Mismatches  246;  Indels   39;  Gaps   11;

Qy      34 VYVGSGGWRFLRIV----CKTARRDLFGLSVLIRVRLELR--RHQRAGHTIPRIFQAVVQ 87
           :| |  |   | ::    |    :|:     |  :  ::|  | :|   ||  :|    :
Db       5 LYTGLAGLLLLPLLLTCCCPYLLQDVRFFLQLANMARQVRSYRQRRPVRTILHVFLEQAR 64

Qy      88 RQPERLALVDAGTGECWTFAQLDAYSNAVAN-LFRQLGFAPGDVVAIFLEGRPEFVGLWL 146
```

```
           : | :  |:     |  |:||:|  || ||  |   ||   || ||:|:   | :| |||
Db     65 KTPHKPFLL--FRDETLTYAQVDRRSNQVARALHDHLGLRQGDCVALFMGNEPAYVWLWL 122

Qy    147 GLAKAGMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIKFCSG 206
          || | |   | || |:| : |  |    ||| |:   |:  || ||   | |
Db    123 GLLKLGCPMACLNYNIRAKSLLHCFQCCGAKVLLASPELHEAVEEVLPTLKK-------- 174

Qy    207 DLGPEGI-------LPDTHLLDPLLKEASTAPLAQIPSKGMDDRLF-----YIYTSGTTG 254
              ||:         :|: :| :| :        ||      : |     |||||||||
Db    175 ----EGVSVFYVSRTSNTNGVDTVLDKVDGVSADPIPESWRSEVTFTTPAVYIYTSGTTG 230

Qy    255 LPKAAIVVHSRYYRMAAFGHHAYRMQAADVLYDCLPLYHSAGNIIGVGQCLIYGLTVVLR 314
          ||||| : | | :   :    :  ::| ||:|  :||||||  :||:  |:: | |  ||
Db    231 LPKAATINHHRLWYGTSLALRS-GIKAHDVIYTTMPLYHSAALMIGLHGCIVVGATFALR 289

Qy    315 KKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRLAVGNGLRPAIWEEF 374
          |||||:||||| ||| ||:|||||: |||   | :  :| |:|::|:|||||  :| ||
Db    290 SKFSASQFWDDCRKYNATVIQYIGELLRYLCNTPQKPNDRDHKVKIALGNGLRGDVWREF 349

Qy    375 TERFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHVYPIRLVKVNEDTMELLRD 434
          :|||   | ||| :|| |    |   |:|: |  : :   |    |:| : :  | :||
Db    350 IKRFGDIHIYEFYASTEGNIGFMNYPRKIGAVGRENYLQKKVVRHELIKYDVEKDEPVRD 409

Qy    435 AQGLCIPCQAGEPGLLVGQINQQDPLRRFDGYV-SESATSKKIAHSVFSKGDSAYLSGDV 493
          | | ||    || |||: :| :  |  | ||   :: | ||    || |||  : |||:
Db    410 ANGYCIKVPKGEVGLLICKITELTP---FFGYAGGKTQTEKKKLRDVFKKGDVYFNSGDL 466

Qy    494 LVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGM 553
          |::|   ::|| || ||||||:||||:|||| ::  :    :| |||| ||| ||: ||
Db    467 LMIDRENFIYFHDRVGDTFRWKGENVATTEVADIVGLVDFVEEVNVYGVPVPGHEGRIGM 526

Qy    554 AAVADPHSL-LDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQKTRLQREGFDPR 612
          |::    :   :   ::| : : |  |:|| |||:   :: ||||| :|  |  |||:|
Db    527 ASIKMKENYEFNGKKLFQHISEYLPSYSRPRFLRIQDTIEITGTFKHRKVTLMEEGFNPS 586

Qy    613 QTSDRLFFLDLKQGHYLPLNEAVYTRI 639
              | |:|:|  :  |:|: | :|  |
Db    587 VIKDTLYFMDDTEKTYVPMTEDIYNAI 613
```

```
RESULT   4
VLCS_HUMAN
ID   VLCS_HUMAN     STANDARD;      PRT;   620 AA.
AC   O14975;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
DE   FATTY-ACID-COA LIGASE).
GN   FACVL1 OR VLACS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RA   Uchiyama A., Aoyama T., Kamijo K., Wakui K., Fukushima Y.,
RA   Shimozawa N., Suzuki Y., Kondo N., Orii T., Hashimoto T.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99216327; PubMed=10198260;
RA   Steinberg S.J., Wang S.J., Kim D.G., Mihalik S.J., Watkins P.A.;
RT   "Human very-long-chain acyl-CoA synthetase: cloning, topography, and
RT   relevance to branched-chain fatty acid metabolism.";
RL   Biochem. Biophys. Res. Commun. 257:615-621(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, PLACENTA AND
CC       PANCREAS.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D88308; BAA23644.1; -.
DR   EMBL; AF096290; AAC64973.1; -.
DR   MIM; 603247; -.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM    107    127       POTENTIAL.
FT   TRANSMEM    262    282       POTENTIAL.
SQ   SEQUENCE   620 AA;  70312 MW;  BFD33E1BB67B3DF8 CRC64;


  Query Match             31.0%;  Score 1044.5;  DB 1;  Length 620;
  Best Local Similarity   40.1%;  Pred. No. 7.1e-74;
  Matches  254;  Conservative   87;  Mismatches  255;  Indels   37;  Gaps   15;

Qy      34 VYVGSGGWRFLRIVCKTARRDLF-GLSVLIRVRLELRR-----HQRAGHTIPRIFQAVVQ 87
           :|    |  || ::        |  :   ::|    ||      :|   || | |    :
Db       5 IYTVLAGLLFLPLLVNLCCPYFFQDIGYFLKVAAVGRRVRSYGQRRPARTILRAFLEKAR 64

Qy      88 RQPERLALVDAGTGECWTFAQLDAYSNAVAN-LFRQLGFAPGDVVAIFLEGRPEFVGLWL 146
           : | :  |:     |  |:||:|  || ||  |   ||   || ||: :   | :| |||
Db      65 QTPHKPFLL--FRDETLTYAQVDRRSNQVARALHDHLGLRQGDCVALLMGNEPAYVWLWL 122

Qy     147 GLAKAGMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKS--LIKFC 204
           || | |   | || |:| : |  |    ||| |:   |: ||| |:   | |    | :
Db     123 GLVKLGCAMACLNYNIRAKSLLHCFQCCGAKVLLVSPELQAAVEEILPSLKKDDVSIYYV 182

Qy     205 SGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKGMDDRLF-----YIYTSGTTGLPKAA 259
           |     :||  |: |    : | || |   ||      :  |     ||||||||||||||
Db     183 SRTSNTDGI--DSFL--DKVDEVSTEP---IPESWRSEVTFSTPALYIYTSGTTGLPKAA 235

Qy     260 IVVHSRYYRMAAFG---HHAYRMQAADVLYDCLPLYHSAGNIIGVGQCLIYGLTVVLRKK 316
           :: | | :    :|        ::| ||:|  || |||| :||:  |:: | |: || |
Db     236 MITHQRIW----YGTGLTFVSGLKADDVIYITLPFYHSAALLIGIHGCIVAGATLALRTK 291

Qy     317 FSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRLAVGNGLRPAIWEEFTE 376
           ||||:|||| ||| ||:|||||: |||   | :  :| |:||||:|||||  :| :| :
Db     292 FSASQFWDDCRKYNVTVIQYIGELLRYLCNSPQKPNDRDHKVRLALGNGLRGDVWRQFVK 351

Qy     377 RFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHVYPIRLVKVNEDTMELLRDAQ 436
           |||   | ||| ||| |    |  |||: |  : :   :   |:| : :  | :||
Db     352 RFGDICIYEFYAATEGNIGFMNYARKVGAVGRVNYLQKKIITYDLIKYDVEKDEPVRDEN 411

Qy     437 GLCIPCQAGEPGLLVGQINQQDPLRRFDGYVSESA-TSKKIAHSVFSKGDSAYLSGDVLV 495
           | |:    || |||| :| |  |   |:||    | | ||    || |||  : |||:|:
Db     412 GYCVRVPKGEVGLLVCKITQLTP---FNGYAGAKAQTEKKKLRDVFKKGDLYFNSGDLLM 468

Qy     496 MDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAA 555
           :|   ::|| || |||||:||||:||||   :  :    :| |||| ||  ||: |||:
Db     469 VDHENFIYFHDRVGDTFRWKGENVATTEVADTVGLVDFVQEVNVYGVHVPDHEGRIGMAS 528

Qy     556 V--ADPHSLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQKTRLQREGFDPRQ 613
           :  : |   |   ::| :   |  |||| |||:   :: ||||| :|  |  |||:|
Db     529 IKMKENHE-FDGKKLFQHIADYLPSYARPRFLRIQDTIEITGTFKHRKMTLVEEGFNPAV 587

Qy     614 TSDRLFFLDLKQGHYLPLNEAVYTRICSGAFAL 646
             | |:|||     |:|: | :|  | :    |
Db     588 IKDALYFLDDTAKMYVPMTEDIYNAISAKTLKL 620


RESULT   5
VLCS_MOUSE
ID   VLCS_MOUSE     STANDARD;      PRT;   620 AA.
```

```
AC   O35488; O70550;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   VERY-LONG-CHAIN ACYL-COA SYNTHETASE (EC 6.2.1.-) (VERY-LONG-CHAIN-
DE   FATTY-ACID-COA LIGASE).
GN   FACVL1 OR VLACS OR VLCS OR SLC27A2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Liver;
RA   Kemp S., Lu J.-F., Smith K.D.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Liver;
RX   MEDLINE=98218572; PubMed=9559670;
RA   Berger J., Truppe C., Neumann H., Forss-Petter S.;
RT   "cDNA cloning and mRNA distribution of a mouse very long-chain acyl-
RT   CoA synthetase.";
RL   FEBS Lett. 425:305-309(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. PEROXISOMAL.
CC   -!- TISSUE SPECIFICITY: STRONG EXPRESSION IN LIVER AND KIDNEY, LOW
CC       EXPRESSION IN BRAIN AND TESTIS, NO EXPRESSION IN SKELETAL MUSCLE
CC       AND SPLEEN.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF033031; AAB87982.1; -.
DR   EMBL; AJ223958; CAA11687.1; -.
DR   MGD; MGI:1347099; Slc27a2.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Fatty acid metabolism; Peroxisome; Transmembrane.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM    107    127       POTENTIAL.
FT   TRANSMEM    262    282       POTENTIAL.
FT   CONFLICT     35     35       Q -> R (IN REF. 2).
SQ   SEQUENCE   620 AA;  70366 MW;  77C98BD0DE3B9FFB CRC64;


  Query Match             30.6%;  Score 1032.5;  DB 1;  Length 620;
  Best Local Similarity   38.7%;  Pred. No. 6.1e-73;
  Matches  241;  Conservative   97;  Mismatches  254;  Indels   31;  Gaps   12;

Qy      34 VYVGSGGWRFLRIV----CKTARRDL---FGLSVLIRVRLELRRHQRAGHTIPRIFQAVV 86
           :| |  |   | ::    |    :|:     |: : | |:   | :|   || | |
Db       5 LYTGLAGLLLLPLLLTCCCPYLLQDVRYFLQLANMAR-RVRSYRQRRPVRTILRAFLEQA 63

Qy      87 QRQPERLALVDAGTGECWTFAQLDAYSNAVAN-LFRQLGFAPGDVVAIFLEGRPEFVGLW 145
           :: | :  |:      |  |:||:|  || ||  |  |||   || ||:|:   | :| :|
Db      64 RKTPHKPFLL--FRDETLTYAQVDRRSNQVARALHDQLGLRQGDCVALFMGNEPAYVWIW 121

Qy     146 LGLAKAGMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIK--F 203
           ||| | |   | || |:| : |  |    ||| |:   ::  || |    | |  :   :
Db     122 LGLLKLGCPMACLNYNIRAKSLLHCFQCCGAKVLLASPDLQEAVEEALPTLKKDAVSVFY 181

Qy     204 CSGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKGMDDRLF-----YIYTSGTTGLPKA 258
           |      |:       | :| :         |     :  |     ||||||||||||
Db     182 VSRTSNTNGV-------DTILDKVDGVSAEPTPESWRSEVTFTTPAVYIYTSGTTGLPKA 234
```

```
Qy     259 AIVVHSRYYRMAAFGHHAYRMQAADVLYDCLPLYHSAGNIIGVGQCLIYGLTVVLRKKFS 318
           | : | | :         :  : | ||:|  :||||||  :||:  |:: | |: || |||
Db     235 ATINHHRLWYGTGLAMSS-GITAQDVIYTTMPLYHSAALMIGLHGCIVVGATLALRSKFS 293

Qy     319 ASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRLAVGNGLRPAIWEEFTERF 378
           ||:||||| ||| ||:|||||: |||   | :  :| |:|: |:||||  :| || :||
Db     294 ASQFWDDCRKYNVTVIQYIGELLRYLCNTPQKPNDRDHKVKKALGNGLRGDVWREFIKRF 353

Qy     379 GVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHVYPIRLVKVNEDTMELLRDAQGL 438
           |   : ||| :|| |    |   |:|: |  : :   |    |:| : :  | :||| |
Db     354 GDIHVYEFYASTEGNIGFVNYPRKIGAVGRANYLQRKVARYELIKYDVEKDEPVRDANGY 413

Qy     439 CIPCQAGEPGLLVGQINQQDPLRRFDGYV-SESATSKKIAHSVFSKGDSAYLSGDVLVMD 497
           ||    || |||| :| |  |   | ||   :: | ||    || |||  : |||:|::|
Db     414 CIKVPKGEVGLLVCKITQLTP---FIGYAGGKTQTEKKKLRDVFKKGDIYFNSGDLLMID 470

Qy     498 ELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVA 557
             ::|| || ||||||:||||:||||  ::  :    :| |||| ||| ||: |||::
Db     471 RENFVYFHDRVGDTFRWKGENVATTEVADIVGLVDFVEEVNVYGVPVPGHEGRIGMASLK 530

Qy     558 DPHSL-LDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQKTRLQREGFDPRQTSD 616
           :   :  ::| : : |  |||| |||:   :: ||||| :|  |  |||:|    |
Db     531 IKENYEEFNGKKLFQHIAEYLPSYARPRFLRIQDTIEITGTFKHRKVTLMEEGFNPTVIKD 590

Qy     617 RLFFLDLKQGHYLPLNEAVYTRI 639
           |:|:|  :  ::|: | :|  | `
Db     591 TLYFMDDAEKTFVPMTENIYNAI 613
```

```
RESULT   6
FAT1_YEAST
ID   FAT1_YEAST     STANDARD;      PRT;   623 AA.
AC   P38225;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE LONG-CHAIN FATTY ACID TRANSPORT PROTEIN.
GN   FAT1 OR YBR041W OR YBR0411.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RA   Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
RA   Vissers S.;
RL   Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RC   STRAIN=W303A;
RX   MEDLINE=97236810; PubMed=9079682;
RA   Faergeman N.J., Dirusso C.C., Elberger A., Knudsen J., Black P.N.;
RT   "Disruption of the Saccharomyces cerevisiae homologue to the murine
RT   fatty acid transport protein impairs uptake and growth on long-chain
RT   fatty acids.";
RL   J. Biol. Chem. 272:8531-8538(1997).
CC   -!- FUNCTION: MAY BE INVOLVED IN LONG-CHAIN FATTY ACIDS UPTAKE, AND
CC       THUS MAY PLAY A PIVOTAL ROLE IN REGULATING THEIR ACCESSIBILITY
CC       PRIOR TO METABOLIC UTILIZATION. MAY PLAY AN IMPORTANT ROLE IN
CC       UPTAKE OF THESE HYDROPHOBIC COMPOUNDS UNDER CONDITIONS WHERE FATTY
CC       ACID SYNTHESIS IS COMPROMISED.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z35910; CAA84983.1; -.
DR   PIR; S45899; S45899.
DR   SGD; S0000245; FAT1.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Lipid transport; Transmembrane.
FT   TRANSMEM      6     26       POTENTIAL.
FT   TRANSMEM     54     71       POTENTIAL.
FT   TRANSMEM    149    169       POTENTIAL.
FT   TRANSMEM    293    313       POTENTIAL.
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    289    289       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    534    534       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    591    591       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   623 AA;  71697 MW;  0AE02700B60C8CFE CRC64;
```

```
  Query Match             20.7%;  Score 699; DB 1;  Length 623;
  Best Local Similarity   33.5%;  Pred. No. 6.8e-47;
  Matches  190;  Conservative   87;  Mismatches  234;  Indels   56;  Gaps   19;

Qy      66 LELRRHQRAGHTIPRIFQAVVQRQPERLA------LVDAGTGECWTFAQLDAYSNAVANL 119
           :::|||:        :|   ||:  : ||      : : |  :  ||  :: | | |  |
Db      69 IDVRRHRFQNW---YLFIKQVQQNGDHLAISYTRPMAEKGEFQLETFTYIETY-NIVLRL 124

Qy     120 FRQLGF----APGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRREPLAFCLGTSG 175
             | |      || |||    :| || ||| |   |   | || | :  ||   |  |
Db     125 SHILHFDYNVQAGDYVAIDCTNKPLFVFLWLSLWNIGAIPAFLNYNTKGTPLVHSLKISN 184

Qy     176 AKALIFGGEMVAAVAEVSGHLGKSLIKFCSGDLGPE---GILPDTHLLDPLLKEASTAPL 232
             :   :    : |    :  :|         |:     | :  |:  ||   |   |
Db     185 ITQVFIDPDASNPIRESEEEIKNAL---------PDVKLNYLEEQDLMHELLNSQSPEFL 235

Qy     233 AQIPSK---GMDD--RLFYIYTSGTTGLPKAAIVVHSRYYRMAAFG----HHAYRMQAAD 283
           |   :   |: |      ||||||||||:||:     :| :: |     |   |
Db     236 QQDNVRTPLGLTDFKPSMLIYTSGTTGLPKSAIM----SWRKSSVGCQVFGHVLHMTNES 291

Qy     284 VLYDCLPLYHSAGNIIGVGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRY 343
           ::  :||:||   ::|    | :|  : |  ||||| ||         | :||:||:|||
Db     292 TVFTAMPLFHSTAALLGACAILSHGGCLALSHKFSASTFWKQVYLTGATHIQYVGEVCRY 351

Qy     344 LLKQPVREAERRHRVRLAVGNGLRPAIWEEFTERFGVRQIGEFYGATE---CNCSIANMD 400
           ||  |: : |: |:|:::| |||||| ||::| :|| :  |||| |||      :   |
Db     352 LLHTPISKYEKMHKVKVAYGNGLRPDIWQDFRKRFNIEVIGEFYAATEAPFATTTFQKGD 411

Qy     401 GKVGSCGFNSRILPHVYPIR--LVKV--NEDTMELLRDAQGLCIPCQAGEPG-LLVGQIN 455
             :|:|     |:      :  ||::  |:|:: : |:::| |     |||| :|:
Db     412 FGIGACRNYGTIIQWFLSFQQTLVRMDPNDDSV-IYRNSKGFCEVAPVGEPGEMLMRIFF 470

Qy     456 QQDPLRRFDGYVSES-ATSKKIAHSVFSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRW 514
           : |   | ||:  :  |  |:   || :||: |  ||:|  ||.|  || || ||||||
Db     471 PKKPETSFQGYLGNAKETKSKVVRDVFRRGDAWYRCGDLLKADEYGLWYFLDRMGDTFRW 530

Qy     515 RGENVSTTEVEGVL--SRLLGQTDVAVYGVAVPGVEGKAGMAAVADPHSLLDPNA----I 568
           : |||||||||  |  |       | | |: ||  ||:|| | :   : ||  |   :
Db     531 KSENVSTTEVEDQLTASNKEQYAQVLVVGIKVPKYEGRAGFAVIKLTDNSLDITAKTKLL 590

Qy     569 YQELQKVLAP-YARPIFLRLLPQVDTT 594
             | ::  | || |:|:: : ::  |
Db     591 NDSLSRLNLPSYAMPLFVKFVDEIKMT 617
```

```
RESULT   7
CAIC_ECOLI
ID   CAIC_ECOLI     STANDARD;      PRT;   522 AA.
AC   P31552;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PROBABLE CROTONOBETAINE/CARNITINE-COA LIGASE (EC 6.3.2.-).
GN   CAIC OR B0037.
```

```
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=O44 K74;
RX   MEDLINE=95115548; PubMed=7815937;
RA   Eichler K., Bourgis F., Buchet A., Kleber H.-P.,
RA   Mandrand-Berthelot M.-A.;
RT   "Molecular characterization of the cai operon necessary for carnitine
RT   metabolism in Escherichia coli.";
RL   Mol. Microbiol. 13:775-786(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92334977; PubMed=1630901;
RA   Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA   Isono K., Mizobuchi K., Nakata A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 0-2.4 min region.";
RL   Nucleic Acids Res. 20:3305-3308(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: COULD CATALYZES THE TRANSFER OF COA TO CROTONOBETAINE
CC       OR CARNITINE.
CC   -!- PATHWAY: CARNITINE METABOLISM.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73904; CAA52113.1; -.
DR   EMBL; D10483; BAA01313.1; -.
DR   EMBL; AE000114; AAC73148.1; -.
DR   PIR; S40558; S40558.
DR   HSSP; P08659; 1LCI.
DR   EcoGene; EG11558; caiC.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM     77     97       POTENTIAL.
FT   TRANSMEM    231    251       POTENTIAL.
FT   VARIANT     103    103       C -> R (IN STRAIN O44 K74).
FT   VARIANT     257    258       RA -> PR (IN STRAIN O44 K74).
FT   VARIANT     291    291       Q -> R (IN STRAIN O44 K74).
FT   VARIANT     333    333       I -> S (IN STRAIN O44 K74).
FT   VARIANT     350    350       V -> A (IN STRAIN O44 K74).
FT   VARIANT     379    379       I -> V (IN STRAIN O44 K74).
FT   VARIANT     393    393       Q -> K (IN STRAIN O44 K74).
FT   VARIANT     401    401       A -> V (IN STRAIN O44 K74).
FT   VARIANT     413    413       R -> C (IN STRAIN O44 K74).
FT   VARIANT     417    417       D -> G (IN STRAIN O44 K74).
FT   VARIANT     447    447       A -> T (IN STRAIN O44 K74).
FT   VARIANT     508    508       L -> LDL (IN STRAIN O44 K74).
SQ   SEQUENCE   522 AA;  59089 MW;  1C54D4003B130671 CRC64;


  Query Match             11.6%;  Score 390.5;  DB 1;  Length 522;
  Best Local Similarity   25.7%;  Pred. No. 6.7e-23;
  Matches  138;  Conservative   84;  Mismatches  251;  Indels   63;  Gaps   16;

Qy      92 RLALV---DAGTGECWTFAQLDAYSNAVANLFRQLGFAPGDVVAIFLEGRPEFVGLWLGL 148
           : ||:     |    ::: :|:   |  ||||  ||   || ||: |:  |||:  | ||
Db      28 KTALICESSGGVVNRYSYLELNQEINRTANLFYTLGIRKGDKVALHLDNCPEFIFCWFGL 87

Qy     149 AKAGMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIKFCSGDL 208
           || |     :|  |  |  |: |  | |  |:   :       ::       |   |  |:
Db      88 AKIGAIMVPINARLLCEESAWILQNSQACLLVTSAQFYPMYQQIQQEDATQLRHICLTDV 147

Qy     209 ---GPEGILPDTHLLDPLLKEASTAPLAQIPSKGMDDRLFYIYTSGTTGLPKAAIVVHSR 265
              :|:   |      ||    | |   |    ||    ::|||||  ||  :: |
Db     148 ALPADDGVSSFTQ-----LKNQQPATLCYAPPLSTDDTAEILFTSGTTSRPKGVVITH-- 200

Qy     266 YYRMAAFGHHA---YRMQAADVLYDCLPLYHSAGNIIGVGQCLIYGLTVVLRKKFSASRF 322
           | :   |:::     ::  ||    :| :|              | | || :|:||  |
Db     201 -YNLRFAGYYSAWQCALRDDDVYLTVMPAFHIDCQCTAAMAAFSAGATFVLVEKYSARAF 259

Qy     323 WDDCIKYNCTVVQYIGEICRYLLKQPVREAERRHRVR-LAVGNGLRPAIWEEFTERFGVR 381
           |    ||  || : |  : | |: ||    :::||:| :     |     : | ||||||
Db     260 WGQVQKYRATVTECIPMMIRTLMVQPPSANDQQHRLREVMFYLNLSEQEKDAFCERFGVR 319

Qy     382 QIGEFYGATECNCSIAN----------MDGKVGSCGFNSRIL-PHVYPIRLVKVNEDTME 430
           :   || ||   |              |:|| | : : |   |  |:   :: |
Db     320 LLTS-YGMTETIVGIIGDRPGDKRRWPSIGRVGFC-YEAEIRDDHNRPLPAGEIGE---- 373

Qy     431 LLRDAQGLCIPCQAGEPGLLVGQINQQDPLRRFDGYVSESATSKKIAHSVFSKGDSAYLS 490
             :||    | ||  : :          : :::  ||:| :      : |    :
Db     374 -------ICI---KGIPGKTIFK----------EYFLNPQATAKVL------EADGWLHT 407

Qy     491 GDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGK 550
           ||    ||  : || ||  :  :  |||||  |:| :::      |: | |:    :  :
Db     408 GDTGYRDEEDFFYFVDRRCNMIKRGGENVSCVELENIIAAHPKIQDIVVVGIK-DSIRDE 466

Qy     551 AGMA-AVADPHSLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGTFKIQKTRLQ 605
           |  | | :    |     ::  :: :| :  | :| :   :    : || :  |:
Db     467 AIKAFVVLNEGETLSEEEFFRFCEQNMAKFKVPSYLEIRKDLPRNCSGKIIRKNLK 522


RESULT   8
PPS2_BACSU
ID   PPS2_BACSU     STANDARD;      PRT;  2560 AA.
AC   P39846;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   PEPTIDE SYNTHETASE 2.
GN   PPSB OR PPS2.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95227362; PubMed=7711903;
RA   Tognoni A., Franchi E., Magistrelli C., Colombo E., Cosmina P.,
RA   Grandi G.;
RT   "A putative new peptide synthase operon in Bacillus subtilis: partial
RT   characterization.";
RL   Microbiology 141:645-648(1995).
CC   -!- COFACTOR: CONTAINS 1 COVALENTLY BOUND PHOSPHOPANTETHEINE
CC       (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; Z34883; CAA84361.1; -.
DR   EMBL; Z99113; CAB13716.1; -.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10971; ppsB.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; DUF4.
DR   InterPro; IPR003880; Phosphopant_attach.
DR   Pfam; PF00501; AMP-binding; 2.
DR   Pfam; PF00668; Condensation; 3.
DR   Pfam; PF00550; pp-binding; 2.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
DR   PROSITE; PS00455; AMP_BINDING; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Multifunctional enzyme; Ligase; Repeat; Phosphopantetheine;
KW   Complete proteome.
FT   DOMAIN      970   1037       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2007   2077       ACYL CARRIER (ACP) 2.
FT   BINDING    2041   2041       PHOSPHOPANTETHEINE (POTENTIAL).
SQ   SEQUENCE   2560 AA;  290161 MW;  2DD2442D11B6E942 CRC64;
```

```
  Query Match             8.9%;  Score 301.5;  DB 1;  Length 2560;
  Best Local Similarity   23.7%;  Pred. No. 5.2e-15;
  Matches  130;  Conservative   82;  Mismatches  251;  Indels   85;  Gaps   13;

Qy      77 TIPRIFQAVVQRQPERLALVDAGTGECWTFAQLDAYSNAVANLFRQLGFAPGDVVAIFLE 136
           |||::|:    : ||  ||      ||||: ||   :| :|:   : |   ||:||: :
Db     470 TIPQLFEEQAHKTPEAAALKMG--NECWTYRQLQVRANQIAHALIEKGVGSGDIVAVMMG 527

Qy     137 GRPEFVGLWLGLAKAGMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHL 196
             |     ||: |||     |: :   | |:| |  | |  |:   ::::
Db     528 RSMEMPAALLGIWKAGGAYMPLDPHFPAERLSFLLKDSQAAQLLIEEDLIS--------- 578

Qy     197 GKSLIKFCSGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKGMDDRLFYIYTSGTTGLP 256
                     ::| ::  : :  | : :   : |:    |  : ||||||| |
Db     579 ----------------LIPPSYEGNTITIEHTESYQTEAPNMPPGDLAYLIYTSGTTGRP 622

Qy     257 KAAIVVHSRYYRMAAFGHHAYRMQAADVLYDCLPLYHSAGNIIGVGQCLIYGLTVVLRKK 316
           |  :| |        :    | |   |:    |  |   | :  :   |: |  |:|
Db     623 KGVLVDHHGIANTLQWRREEYSMTEQDISLH-LFSYVFDGCVTSLFTPLLSGACVLLTTD 681

Qy     317 FSASRFWDDCIKYNCTVVQY-------IGEICRYLLKQPVREAERRHRVRLAVGNGLRPA 369
             |    | :    : :|        :  : | ||:    :  :  |:    |  : |
Db     682 DEAK----DVLALKRKIARYKVSHMIIVPSLYRVLLEVMTADDAKSLRIVTFAGEAVTPD 737

Qy     370 IWEEFTERFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHV---------YPIR 420
           : |          ::   || ||   |:|            : || |:          :|||
Db     738 LLELNQIICPSAELANEYGPTE--NSVA-----------TTILRHLNKKERITIGHPIR 783

Qy     421 LVKV---NEDTMELLRDAQGLCIPCQAGEPGLLVGQINQQDPLRRFDGYVSESATSKKIA 477
             ||   : : |: :  |  |||       ||  |   ||:             | |  :
Db     784 NTKVFVLHGNQMQPIGAAGELCI----SGAGLARGYYKQQE-----------LTQKAFS 827

Qy     478 HSVFSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDV 537
             | :|:  | :||       | : :  |  |  : ||  :   |:| || :  |  :
Db     828 DHPFLEGERLYRTGDAGRFLPDGTIEYIGRFDDQVKIRGYRIELREIETVLRQAPGVKEA 887

Qy     538 AVYGVAVPGVEGKAGMAAVADP-HSLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTGT 596
           ||    |   | :     | :  :|| |   :|| |   |  |  |  :  : |:  | :
Db     888 AVLARDVSAEEKELVAYIVPEKGNSLPD---LYQHLAGTLPSYMIPASIINISQMPLTSS 944

Qy     597 FKIQKTRL 604
             |: :  |
Db     945 GKLDRFAL 952
```

```
RESULT   9
4CL1_PETCR
ID   4CL1_PETCR     STANDARD;      PRT;   544 AA.
AC   P14912;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA
DE   SYNTHASE 1).
GN   4CL1 OR 4CL-1.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89005119; PubMed=3169018;
RA   Lozoya E., Hoffmann H., Douglas C., Schulz W., Scheel D.,
RA   Hahlbrock K.;
RT   "Primary structures and catalytic properties of isoenzymes encoded by
RT   the two 4-coumarate:CoA ligase genes in parsley.";
RL   Eur. J. Biochem. 176:661-667(1988).
RN   [2]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Douglas C., Hoffmann H., Schulz W., Hahlbrock K.;
RT   "Structure and elicitor or U.V.-light-stimulated expression of two
RT   4-coumarate:CoA ligase genes in parsley.";
RL   EMBO J. 6:1189-1195(1987).
CC   -!- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
CC       4-COUMAROYL-COA.
CC   -!- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC       METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC   -!- INDUCTION: TRANSCRIPTIONALLY ACTIVATED BY BOTH FUNGAL ELICITOR AND
CC       ULTRAVIOLET IRRADIATION.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X13324; CAA31696.1; -.
DR   EMBL; X05350; CAA28959.1; -.
DR   PIR; S01667; S01667.
DR   HSSP; P08659; 1LCI.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Phenylpropanoid metabolism; Multigene family.
SQ   SEQUENCE   544 AA;  59825 MW;  22BBAD78F255D0C8 CRC64;
```

```
  Query Match             8.9%;  Score 298.5;  DB 1;  Length 544;
  Best Local Similarity   23.9%;  Pred. No. 1.1e-15;
  Matches  128;  Conservative   95;  Mismatches  231;  Indels   81;  Gaps   18;

Qy      48 CKTARRDLFGLSVL--IRVRLELRRHQRAGHTIPRIFQAVVQRQPERLALVDAGTGECWT 105
             |   : ||   | |  | :   |  |       | ::         ::  |::  ||| :|
Db       4 CVAPKEDLIFRSKLPDIYIPKHLPLHTYCFENISKV--------GDKSCLINGATGETFT 55

Qy     106 FAQLDAYSNAVANLFRQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRRE 165
           ::|::  |  ||:   :||   || : : |   ||:   :|| :  |  : : |
Db      56 YSQVELLSRKVASGLNKLGIQQGDTIMLLLPNSPEYFFAFLGASYRGAISTMANPFFTSA 115

Qy     166 PLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIKFCSGDLGPEGILPDTHLLDPLLK 225
             :   |  | || :|        | :|  :  :  |:    |  |:  |  : |::
Db     116 EVIKQLKASQAKLII---TQACYVDKVKDYAAEKNIQIICIDDAPQDCLHFSKLME---A 169
```

```
Qy    226 EASTAPLAQIPSKGMDDRLFYIYTSGTTGLPKAAIVVH-----SRYYRMAAFGHHAYRMQ 280
Db    170 DESEMPEVVINS---DDVVALPYSSGTTGLPKGVMLTHKGLVTSVAQQVDGDNPNLY-MH 225

Qy    281 AADVLYDCLPLYH--SAGNIIGVGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIG 338
Db    226 SEDVMICILPLFHIYSLNAVLCCG--LRAGVTILIMQKFDIVPFLELIQKYKVTIGPFVP 283

Qy    339 EICRYLLKQPVREAERRHRVRLA------VGNGLRPAIWEEFTERFGVRQIGEFYGATEC 392
Db    284 PIVLAIAKSPVVDKYDLSSVRTVMSGAAPLGKELEDAVRAKFPN----AKLGQGYGMTEA 339

Qy    393 NCSIA--------NMDGKVGSCGFNSRILPHVYPIRLVKVNEDTMELLRDAQGLCIP-CQ 443
Db    340 GPVLAMCLAFAKEPYEIKSGACG-------------TVVRNAEMKIVDPETNASLPRNQ 385

Qy    444 AGEPGLLVGQINQQDPLRRFDGYVSESATSKKIAHSVFSKGDSAYL-SGDVLVMDELGYM 502
Db    386 RGEICIRGDQI--------MKGYLNDPESTRT------TIDEEGWLHTGDIGFIDDDDEL 431

Qy    503 YFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVA 557
Db    432 FIVDRLKEIIKYKGFQVAPAELEALLLTHPTISDAAV----VPMIDEKAGEVPVA 482


RESULT 10
4CL2_PETCR
ID   4CL2_PETCR     STANDARD;      PRT;   544 AA.
AC   P14913;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA
DE   SYNTHASE 1).
GN   4CL2 OR 4CL-2.
OS   Petroselinum crispum (Parsley) (Petroselinum hortense).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.
OX   NCBI_TaxID=4043;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89005119; PubMed=3169018;
RA   Lozoya E., Hoffmann H., Douglas C., Schulz W., Scheel D.,
RA   Hahlbrock K.;
RT   "Primary structures and catalytic properties of isoenzymes encoded by
RT   the two 4-coumarate:CoA ligase genes in parsley.";
RL   Eur. J. Biochem. 176:661-667(1988).
RN   [2]
RP   SEQUENCE OF 1-8 FROM N.A.
RA   Douglas C., Hoffmann H., Schulz W., Hahlbrock K.;
RT   "Structure and elicitor or U.V.-light-stimulated expression of two
RT   4-coumarate:CoA ligase genes in parsley.";
RL   EMBO J. 6:1189-1195(1987).
CC   -!- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
CC       4-COUMAROYL-COA.
CC   -!- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC       METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC   -!- INDUCTION: TRANSCRIPTIONALLY ACTIVATED BY BOTH FUNGAL ELICITOR AND
CC       ULTRAVIOLET IRRADIATION.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X13325; CAA31697.1; -.
DR   EMBL; X05351; CAA28960.1; -.
DR   PIR; S01667; S01667.
DR   PIR; S15695; S15695.
DR   HSSP; P08659; 1LCI.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Phenylpropanoid metabolism; Multigene family.
SQ   SEQUENCE   544 AA;  59783 MW;  B477965C68F8C534 CRC64;


  Query Match             8.8%;  Score 297.5;  DB 1;  Length 544;
  Best Local Similarity   23.9%;  Pred. No. 1.3e-15;
  Matches  128;  Conservative   95;  Mismatches  231;  Indels   81;  Gaps   18;

Qy     48 CKTARRDLFGLSVL--IRVRLELRRHQRAGHTIPRIFQAVVQRQPERLALVDAGTGECWT 105
Db      4 CVAPKEDLIFRSKLPDIYIPKHLPLHTYCFENISKV--------GDKSCLINGATGETFT 55

Qy    106 FAQLDAYSNAVANLFRQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRRE 165
Db     56 YSQVELLSRKVASGLNKLGIQQGDTIMLLLPNSPEYFFAFLGASYRGAISTMANPFFTSA 115

Qy    166 PLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIKFCSGDLGPEGILPDTHLLDPLLK 225
Db    116 EVIKQLKASLAKLII---TQACYVDKVKDYAAEKNIQIICIDDAPQDCLHFSKLME---A 169

Qy    226 EASTAPLAQIPSKGMDDRLFYIYTSGTTGLPKAAIVVH-----SRYYRMAAFGHHAYRMQ 280
Db    170 DESEMPEVVIDS---DDVVALPYSSGTTGLPKGVMLTHKGLVTSVAQQVDGDNPNLY-MH 225

Qy    281 AADVLYDCLPLYH--SAGNIIGVGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIG 338
Db    226 SEDVMICILPLFHIYSLNAVLCCG--LRAGVTILIMQKFDIVPFLELIQKYKVTIGPFVP 283

Qy    339 EICRYLLKQPVREAERRHRVRLA------VGNGLRPAIWEEFTERFGVRQIGEFYGATEC 392
Db    284 PIVLAIAKSPVVDKYDLSSVRTVMSGAAPLGKELEDAVRAKFPN----AKLGQGYGMTEA 339

Qy    393 NCSIA--------NMDGKVGSCGFNSRILPHVYPIRLVKVNEDTMELLRDAQGLCIP-CQ 443
Db    340 GPVLAMCLAFAKEPYEIKSGACG-------------TVVRNAEMKIVDPETNASLPRNQ 385

Qy    444 AGEPGLLVGQINQQDPLRRFDGYVSESATSKKIAHSVFSKGDSAYL-SGDVLVMDELGYM 502
Db    386 RGEICIRGDQI--------MKGYLNDPESTRT------TIDEEGWLHTGDIGFIDDDDEL 431

Qy    503 YFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVA 557
Db    432 FIVDRLKEIIKYKGFQVAPAELEALLLTHPTISDAAV----VPMIDEKAGEVPVA 482


RESULT 11
SRF1_BACSU
ID   SRF1_BACSU     STANDARD;      PRT;  3587 AA.
AC   P27206;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SURFACTIN SYNTHETASE SUBUNIT 1.
GN   SRFAA OR SRFA1 OR SRFA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA   Zuber P., Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
RL   Nucleic Acids Res. 21:93-97(1993).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=93360813; PubMed=8355609;
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821-831(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [4]
RP   SEQUENCE OF 1-460 FROM N.A.
RX   MEDLINE=91154134; PubMed=1847909;
RA   Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA   Zuber P.;
RT   "srfA is an operon required for surfactin production, competence
RT   development, and efficient sporulation in Bacillus subtilis.";
RL   J. Bacteriol. 173:1770-1778(1991).
RN   [5]
RP   SEQUENCE OF 1-38 FROM N.A.
RX   MEDLINE=91358326; PubMed=1715856;
RA   Nakano M.M., Xia L., Zuber P.;
RT   "Transcription initiation region of the srfA operon, which is
RT   controlled by the comP-comA signal transduction system in Bacillus
RT   subtilis.";
RL   J. Bacteriol. 173:5487-5493(1991).
RN   [6]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=95219080; PubMed=7704255;
RA   Fujishima Y., Yamane K.;
RT   "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)
RT   of srfAA of the Bacillus subtilis chromosome.";
RL   Microbiology 141:277-279(1995).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC       ANTIBIOTIC SURFACTIN.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D13262; BAA02522.1; -.
DR   EMBL; X70356; CAA49816.1; -.
DR   EMBL; D50453; BAA08982.1; -.
DR   EMBL; M59939; AAA22815.1; -.
DR   EMBL; M64702; AAA22816.1; -.
DR   EMBL; D30762; BAA21034.1; -.
DR   EMBL; Z99105; CAB12142.1; -.
DR   PIR; S35517; S35517.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10168; srfAA.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; DUF4.
DR   InterPro; IPR003880; Phosphopant_attach.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
DR   Pfam; PF00550; pp-binding; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW   Multifunctional enzyme; Repeat; Complete proteome.
FT   REPEAT        ?    1047       DOMAIN 1 (GLU-ACTIVATING).
FT   REPEAT        ?    2086       DOMAIN 2 (LEU-ACTIVATING).
FT   REPEAT        ?    3114       DOMAIN 3 (D-LEU-ACTIVATING).
FT   DOMAIN      976    1043       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2015    2082       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3043    3109       ACYL CARRIER (ACP) 3.
FT   BINDING    1006    1006       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2045    2045       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3073    3073       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT    146     146       I -> M (IN REF. 2 AND 3).
FT   CONFLICT    151     151       L -> V (IN REF. 2 AND 3).
FT   CONFLICT    165     165       A -> G (IN REF. 2, 3 AND 4).
FT   CONFLICT    281     281       T -> Q (IN REF. 2 AND 3).
FT   CONFLICT    460     460       D -> T (IN REF. 3 AND 3).
FT   CONFLICT    540     540       P -> A (IN REF. 2 AND 3).
FT   CONFLICT    562     562       I -> Y (IN REF. 2 AND 3).
FT   CONFLICT    639     640       GS -> PT (IN REF. 2 AND 3).
FT   CONFLICT    644     644       R -> P (IN REF. 2 AND 3).
FT   CONFLICT    647     647       L -> F (IN REF. 2 AND 3).
FT   CONFLICT    649     649       A -> P (IN REF. 2 AND 3).
FT   CONFLICT   1026    1026       L -> Q (IN REF. 2 AND 3).
FT   CONFLICT   1065    1116       SLAQQRTYIVSQFEDAGVGYNMPAAAILEGPLDIQKLERAF
FT                                 QGLIRRHESLR -> HWHSSEHISSASSRMRESAIHASSSN
FT                                 SEGFRYSKAGARISGINPTPRVIE (IN REF. 2 AND
FT                                 3).
FT   CONFLICT   1132    1133       DS -> VC (IN REF. 2 AND 3).
FT   CONFLICT   1164    1164       R -> A (IN REF. 2 AND 3).
FT   CONFLICT   1458    1458       V -> L (IN REF. 2 AND 3).
FT   CONFLICT   1850    1850       S -> T (IN REF. 2 AND 3).
FT   CONFLICT   1894    1894       D -> T (IN REF. 2 AND 3).
FT   CONFLICT   1911    1911       R -> A (IN REF. 2 AND 3).
FT   CONFLICT   1973    1980       LRCLSKWT -> PAVFIQMD (IN REF. 2 AND 3).
FT   CONFLICT   2052    2052       L -> P (IN REF. 2 AND 3).
FT   CONFLICT   2218    2218       H -> Q (IN REF. 2 AND 3).
FT   CONFLICT   2265    2266       QQ -> HR (IN REF. 2 AND 3).
FT   CONFLICT   2291    2291       V -> L (IN REF. 2 AND 3).
FT   CONFLICT   2349    2349       S -> C (IN REF. 2 AND 3).
FT   CONFLICT   2428    2428       R -> S (IN REF. 2 AND 3).
FT   CONFLICT   2611    2612       RC -> AV (IN REF. 2 AND 3).
FT   CONFLICT   2685    2685       E -> ENPE (IN REF. 2 AND 3).
FT   CONFLICT   2755    2755       T -> S (IN REF. 2 AND 3).
FT   CONFLICT   2896    2897       SP -> TA (IN REF. 2 AND 3).
FT   CONFLICT   3025    3025       P -> N (IN REF. 2 AND 3).
FT   CONFLICT   3096    3096       F -> N (IN REF. 2 AND 3).
FT   CONFLICT   3271    3271       A -> S (IN REF. 2 AND 3).
FT   CONFLICT   3316    3316       R -> S (IN REF. 2 AND 3).
FT   CONFLICT   3451    3451       Y -> S (IN REF. 2 AND 3).
FT   CONFLICT   3483    3484       DE -> HQ (IN REF. 2 AND 3).
FT   CONFLICT   3487    3490       DAGL -> HPPF (IN REF. 2 AND 3).
FT   CONFLICT   3493    3493       R -> T (IN REF. 2 AND 3).
FT   CONFLICT   3495    3495       E -> H (IN REF. 2 AND 3).
FT   CONFLICT   3499    3500       GQ -> PH (IN REF. 2 AND 3).
FT   CONFLICT   3507    3507       R -> E (IN REF. 2 AND 3).
SQ   SEQUENCE   3587 AA;  402424 MW;  A1E2DABFF93EDE3A CRC64;


  Query Match             8.7%;  Score 293.5;  DB 1;  Length 3587;
  Best Local Similarity   22.1%;  Pred. No. 3.5e-14;
  Matches  125;  Conservative   90;  Mismatches  226;  Indels  125;  Gaps   18;

Qy      77 TIPRIFQAVVQRQPERLALVDAGTGECWTFAQLDAYSNAVANLFRQLGFAPGDVVAIFLE 136
           |: ::|:   ||  :| |:     |: ||: :|:| :| :| :    | :| | | :  :
Db    2546 TVHQLFEETAQRHKDRPAV--TYNGQSWTYGELNAKANRLARILMDCGISPDDRVGVLTK 2603

Qy     137 GRPEFVGLWLGLAKAGMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHL 196
```

```
Db   2604 PSLEMSARCLGVLKAGAAFVPIDPDYPDQRIEYILQDSGAKLL----------------- 2646

Qy    197 GKSLIKFCSGDLGPEGI-LPDTHLLDPLLKEASTAPLAQIPSKGMDDR-----------L 244

Db   2647 -----------LKQEGISVPDSYTGDVILLDGSRTILS-LPLDENDEENPETAVTAENLA 2694

Qy    245 FYIYTSGTTGLPKAAIVVHSRYYRMAAFGHHAYRMQAAD------------VLYDCLPLY 292

Db   2695 YMIYTSGTTGQPKGVMVEHHALVNLCFWHHDAFSMTAEDRSAKYAGFGFDASIWEMFPTW 2754

Qy    293 HSAGNIIGVGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREA 352

Db   2755 TIGAELHVIEEAIRLDIVRL-----------NDYFETNGVTITFL----------PTQLA 2793

Qy    353 ER-------RHRVRLAVGNGLRPAIWEEFTERFGVRQIGEFYGATECNCSIANMDGKVGS 405

Db   2794 EQFMELENTSLRVLLTGGDKLKRAVKKPYT-------LVNNYGPTE-NTVVAT------- 2838

Qy    406 CGFNSRILPHVYPIRLVKVNEDTMELLRDAQGLCIP-------CQAGEPGLLVGQINQQD 458

Db   2839 ---SAEIHPEEGSLSIGRAIANTRVYILGEGNQVQPEGVAGELCVAGR-GLARGYLNRED 2894

Qy    459 PLRRFDGYVSESATSKKIAHSVFSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGEN 518

Db   2895 -----------ESPKRFVADPFVPGERMYRTGDLVKWTGGGIEYI-GRIDQQVKVRGYR 2941

Qy    519 VSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVADPHSLLDPNAIYQELQKVLAP 578

Db   2942 IELSEIEVQLAQLSEVQDAAVTAVKDKG--GNTAIAAYVTPES-ADIEALKSALKETLPD 2998

Qy    579 YARPIFLRLLPQVDTTGTFKIQKTRL 604

Db   2999 YMIPAFWVTLNELPVTANGKVDRKAL 3024


RESULT 12
4CL1_TOBAC
ID   4CL1_TOBAC     STANDARD;      PRT;   547 AA.
AC   O24145;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA
DE   SYNTHASE 1).
GN   4CL1.
OS   Nicotiana tabacum (Common tobacco).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX   NCBI_TaxID=4097;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96416441; PubMed=8819324;
RA   Lee D., Douglas C.J.;
RT   "Two divergent members of a tobacco 4-coumarate:coenzyme A ligase
RT   (4CL) gene family. cDNA structure, gene inheritance and expression,
RT   and properties of recombinant proteins.";
RL   Plant Physiol. 112:193-205(1996).
CC   -!- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
CC       4-COUMAROYL-COA.
CC   -!- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC       METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U50845; AAB18637.1; -.
DR   HSSP; P08659; 1LCI.
DR   Mendel; 24477; Nicta;1179;24477.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Phenylpropanoid metabolism; Multigene family.
SQ   SEQUENCE   547 AA;  59842 MW;  154DF6D684E3F51B CRC64;


  Query Match             8.6%;  Score 290.5;  DB 1;  Length 547;
  Best Local Similarity   24.7%;  Pred. No. 4.6e-15;
  Matches  121;  Conservative   82;  Mismatches  215;  Indels   71;  Gaps   17;

Qy     92 RLALVDAGTGECWTFAQLDAYSNAVANLFRQLGFAPGDVVAIFLEGRPEFVGLWLGLAKA 151

Db     46 RPCLINGANDQIYTYAEVELTCRKVAVGLNKLGIQQKDTIMILLPNSPEFVFAFMGASYL 105

Qy    152 GMEAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIKFCSGDLGPE 211

Db    106 GAISTMANPLFTPAEVVKQAKASSAKIII---TQSCFVGKVKDYASENDVKVICIDSAPE 162

Qy    212 GILPDTHLLDPLLKEASTAPLAQIPSKGMDDRLFYIYTSGTTGLPKAAIVVHSRYYRMAA 271

Db    163 GCL---HFSELTQSDEHEIPEVKIQP---DDVVALPYSSGTTGLPKGVMLTHKGLVTSVA 216

Qy    272 ---FGHHA-YRMQAADVLYDCLPLYH--SAGNIIGVGQCLIYGLTVVLRKKFSASRFWDD 325

Db    217 QQVDGENANLYMHSEDVLMCVLPLFHIYSLNSILLCG--LRVGAAILIMQKFDIAPFLEL 274

Qy    326 CIKYNCTVVQYIGEICRYLLKQPVREAERRHRVRLA------VGNGLRPAIWEEFTERFG 379

Db    275 IQKYKVSIGPFVPPIVLAIAKSPIVDSYDLSSVRTVMSGAAPLGKELEDAVRTKFPN--- 331

Qy    380 VRQIGEFYGATECNCSIA--------NMDGKVGSCGFNSRILPHVYPIRLVKVNEDTMEL 431

Db    332 -AKLGQGYGMTEAGPVLAMCLAFAKEPFDIKSGACG--------------TVVRNAEMKI 376

Qy    432 LRDAQGLCIP-CQAGEPGLLVGQINQQDPLRRFDGYVSE-SATSKKIAHSVFSKGDSAYL 489

Db    377 VDPDTGCSLPRNQPGEICIRGDQI--------MKGYLNDPEATTRTI-------DKEGWL 421

Qy    490 -SGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVE 548

Db    422 HTGDIGFIDEDDELFIVDRLKELIKYKGFQVAPAEIEALLLNHPNISDAAV----VPMKD 477

Qy    549 GKAGMAAVA 557

Db    478 EQAGEVPVA 486


RESULT 13
SRF2_BACSU
ID   SRF2_BACSU     STANDARD;      PRT;  3587 AA.
AC   Q04747;
DT   01-FEB-1995 (Rel. 31, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   SURFACTIN SYNTHETASE SUBUNIT 2.
GN   SRFAB OR SRFA2 OR COML.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE OF 1-3077 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=93181186; PubMed=8441623;
RA   Fuma S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA   Zuber P., Yamane K.;
RT   "Nucleotide sequence of 5' portion of srfA that contains the region
RT   required for competence establishment in Bacillus subtilis.";
```

```
RL   Nucleic Acids Res. 21:93-97(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / JH642;
RX   MEDLINE=93360813; PubMed=8355609;
RA   Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA   Venema G., van Sinderen D.;
RT   "Sequence and analysis of the genetic locus responsible for surfactin
RT   synthesis in Bacillus subtilis.";
RL   Mol. Microbiol. 8:821-831(1993).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97124189; PubMed=8969502;
RA   Yamane K., Kumano M., Kurita K.;
RT   "The 25 degrees-36 degrees region of the Bacillus subtilis
RT   chromosome: determination of the sequence of a 146 kb segment and
RT   identification of 113 genes.";
RL   Microbiology 142:3047-3056(1996).
RN   [4]
RP   SEQUENCE OF 514-800 FROM N.A.
RC   STRAIN=ATCC 21332;
RX   MEDLINE=92290255; PubMed=1601288;
RA   Borchert S., Patil S.S., Marahiel M.A.;
RT   "Identification of putative multifunctional peptide synthetase genes
RT   using highly conserved oligonucleotide sequences derived from known
RT   synthetases.";
RL   FEMS Microbiol. Lett. 71:175-180(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC       ANTIBIOTIC SURFACTIN.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13262; BAA02523.1; -.
DR   EMBL; X70356; CAA49817.1; -.
DR   EMBL; D50453; BAA08983.1; -.
DR   EMBL; X65835; CAA46678.1; -.
DR   EMBL; Z99105; CAB12143.1; -.
DR   HSSP; P14687; 1AMU.
DR   SubtiList; BG10169; srfAB.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; DUF4.
DR   InterPro; IPR003880; Phosphopant_attach.
DR   Pfam; PF00501; AMP-binding; 3.
DR   Pfam; PF00668; Condensation; 4.
DR   Pfam; PF00550; pp-binding; 3.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00455; AMP_BINDING; 3.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Ligase; Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
KW   Multifunctional enzyme; Repeat; Complete proteome.
FT   REPEAT        ?   1040       DOMAIN 1 (VAL-ACTIVATING).
FT   REPEAT        ?   2096       DOMAIN 2 (ASP-ACTIVATING).
FT   REPEAT        ?   3114       DOMAIN 3 (D-LEU-ACTIVATING).
FT   DOMAIN      970   1036       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2015   2082       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3043   3109       ACYL CARRIER (ACP) 3.
FT   BINDING     999    999       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    2045   2045       PHOSPHOPANTETHEINE (POTENTIAL).
FT   BINDING    3073   3073       PHOSPHOPANTETHEINE (POTENTIAL).
FT   CONFLICT     33     33       S -> F (IN REF. 1).
FT   CONFLICT     42     42       A -> G (IN REF. 1).
FT   CONFLICT    110    110       Q -> D (IN REF. 1).
FT   CONFLICT    113    115       RQA -> AQG (IN REF. 1).
FT   CONFLICT    139    139       A -> V (IN REF. 1).
FT   CONFLICT    259    259       L -> W (IN REF. 1).
FT   CONFLICT    309    309       R -> A (IN REF. 1).
FT   CONFLICT    478    480       TPA -> SRP (IN REF. 1).
FT   CONFLICT    596    596       MISSING (IN REF. 4).
FT   CONFLICT    648    648       A -> R (IN REF. 1).
FT   CONFLICT    680    682       RHV -> ETL (IN REF. 1).
FT   CONFLICT    694    698       EQSIT -> DKRIS (IN REF. 4).
FT   CONFLICT    788    788       M -> L (IN REF. 4).
FT   CONFLICT    939    940       PL -> LV (IN REF. 1).
FT   CONFLICT   1038   1038       N -> I (IN REF. 1).
FT   CONFLICT   1133   1133       H -> Q (IN REF. 1).
FT   CONFLICT   1310   1310       V -> C (IN REF. 1).
FT   CONFLICT   1333   1333       G -> V (IN REF. 1).
FT   CONFLICT   1384   1384       P -> R (IN REF. 1).
FT   CONFLICT   1582   1582       G -> E (IN REF. 1).
FT   CONFLICT   1677   1682       KRRADG -> E (IN REF. 1).
FT   CONFLICT   1700   1700       C -> S (IN REF. 1).
FT   CONFLICT   1755   1755       F -> K (IN REF. 1).
FT   CONFLICT   1787   1787       T -> S (IN REF. 1).
FT   CONFLICT   1801   1822       GAIAGRVDLYEPDAFAKRPTIG ->
FT                                APSPGGLICMSRCICETPDNR (IN REF. 1).
FT   CONFLICT   1915   1916       LG -> PK (IN REF. 1).
FT   CONFLICT   2075   2075       R -> C (IN REF. 1).
FT   CONFLICT   2079   2079       A -> V (IN REF. 1).
FT   CONFLICT   2141   2145       ARLTP -> LRDSLN (IN REF. 1).
FT   CONFLICT   2445   2445       E -> Q (IN REF. 1).
FT   CONFLICT   2485   2489       ATDLF -> RQICS (IN REF. 1).
FT   CONFLICT   2546   2567       TVHQLFEETVQRHKDRPAVTYN -> DGCISYSKRLSSATK
FT                                TARLSHTN (IN REF. 1).
FT   CONFLICT   2608   2615       MSAAVLGV -> KCPPRCSAS (IN REF. 1).
FT   CONFLICT   2644   2645       KL -> NV (IN REF. 1).
FT   CONFLICT   2713   2713       D -> H (IN REF. 1).
FT   CONFLICT   2723   2723       D -> H (IN REF. 1).
FT   CONFLICT   2876   2881       GELCVA -> RALRG (IN REF. 1).
FT   CONFLICT   2899   2900       RF -> L (IN REF. 1).
FT   CONFLICT   2958   2960       EDR -> QDA (IN REF. 1).
FT   CONFLICT   2964   2964       R -> A (IN REF. 1).
SQ   SEQUENCE   3587 AA;  401248 MW;  6B0B05A9FF32054D CRC64;
```

```
  Query Match             8.5%;  Score 287.5;  DB 1;  Length 3587;
  Best Local Similarity   21.5%;  Pred. No. 1e-13;
  Matches  131;  Conservative   96;  Mismatches  249;  Indels  133;  Gaps   19;

Qy      42 RFLRIVCKTARRDLFGLSVLIRVRLELRRHQRAGHTIP--------RIFQAVVQRQPERL 93
           | |: |     : |  | ::     :       |  :|       ::|:  |||  :|
Db    2503 RLLKAVADNPNQPLSSLLLVTETEKQALLEAWKGKALPVPTDKTVHQLFEETVQRHKDRP 2562

Qy      94 ALVDAGTGECWTFAQLDAYSNAVANLFRQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGM 153
           |:     |: ||: :|:| :| :| :    | :| | | :  :   |     ||: |||
Db    2563 AV--TYNGQSWTYGELNAKANRLARILMDCGISPDDRVGVLTKPSLEMSAAVLGVLKAGA 2620

Qy     154 EAALLNVNLRREPLAFCLGTSGAKALIFGGEMVAAVAEVSGHLGKSLIKFCSGDLGPEGI 213
           :: :   : : : |  |||| |                                  |  |||
Db    2621 AFVPIDPDYPDQRIEYILQDSGAKLL---------------------------LKQEGI 2652

Qy     214 -LPDTHLLDPLLKEASTAPLAQIPSKGMDD----------RLFYIYTSGTTGLPKAAIV 261
            :||::  | :| :: |   |: :|    |:          : |||||||| ||  :|
Db    2653 SVPDSYTGDVILLDGSRTILS-LPLDENDEGNPETAVTAENLAYMIYTSGTTGQPKGVMV 2711

Qy     262 VHSRYYRMAAFGHHAYRMQAAD----------VLYDCLPLYHSAGNIIGVGQCLIYGL 309
                :  : | |: | | |              :::  | :     :  : : :   :
Db    2712 EDHALVNLCFWDHDAFSMTAEDRSAKYAGFGFDASIWEMFPTWTIGAELHVIDEAIRLDI 2771

Qy     310 TVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVREAER-------RHRVRLAV 362
             :          :|  : |   : ::          | : ||:         || |
Db    2772 VRL----------NDYFETNGVTITFL----------PTQLAEQFMELENTSLRVLLTG 2810
```

```
Qy     363 GNGLRPAIWEEFTERFGVRQIGEFYGATECNCSIANMDGKVGSCGFNSRILPHVYPIRLV 422
           |: |: |: : :|       :   || || |  :|           :: | |    : :
Db    2811 GDKLKRAVKKPYT-------LVNNYGPTE-NTVVAT----------SAEIHPEEGSLSIG 2852

Qy     423 KVNEDTMELLRDAQGLCIP-------CQAGEPGLLVGQINQQDPLRRFDGYVSESATSKK 475
           :   :|   :        |       | ||  ||  | :|::|             |:|:
Db    2853 RAIANTRVYILGEGNQVQPEGVAGELCVAGR-GLARGYLNRED------------ETAKR 2899

Qy     476 IAHSVFSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRGENVSTTEVEGVLSRLLGQT 535
              |  |:  | :||::     |  |    |     : ||  :  :|:|  |::|
Db    2900 FVADPFVPGERMYRTGDLVKWVNGGIEYI-GRIDQQVKVRGYRIELSEIEVQLAQLSEVE 2958

Qy     536 DVAVYGVAVPGVEGKAGMAAVADPHSLLDPNAIYQELQKVLAPYARPIFLRLLPQVDTTG 595
           | ||  |   |  |   :||   | :  |  |:   |:: |  |  | |   | ::  |
Db    2959 DRAVTRVKDKG--GNTAIAAYVTPET-ADIEALKSTLKETLPDYMIPAFWVTLNELPVTA 3015

Qy     596 TFKIQKTRL 604
             |: :  |
Db    3016 NGKVDRKAL 3024


RESULT 14
4CL3_ARATH
ID   4CL3_ARATH     STANDARD;      PRT;   561 AA.
AC   Q9S777;
DT   20-AUG-2001 (Rel. 40, Created)
DT   20-AUG-2001 (Rel. 40, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   4-COUMARATE--COA LIGASE 3 (EC 6.2.1.12) (4CL 3) (4-COUMAROYL-COA
DE   SYNTHASE 3).
GN   4CL3 OR AT1G65060 OR F16G16.6.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=99348176; PubMed=10417722;
RA   Ehlting J., Buettner D., Wang Q., Douglas C.J., Somssich I.E.,
RA   Kombrink E.;
RT   "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana
RT   represent two evolutionarily divergent classes in angiosperms.";
RL   Plant J. 19:9-20(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
CC   -!- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
CC       4-COUMAROYL-COA.
CC   -!- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC       METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF106087; AAD47194.1; -.
DR   EMBL; AF106088; AAD47195.1; -.
DR   EMBL; AC009360; AAF06039.1; -.
DR   HSSP; P08659; 1LCI.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Phenylpropanoid metabolism; Multigene family.
SQ   SEQUENCE   561 AA;  61310 MW;  C2EFF1C36B33F6DC CRC64;


  Query Match             8.3%;  Score 281.5;  DB 1;  Length 561;
  Best Local Similarity   24.6%;  Pred. No. 2.4e-14;
  Matches  128;  Conservative   83;  Mismatches  222;  Indels   87;  Gaps   20;

Qy      79 PRIFQA------VVQRQP------ERLA-------LVDAGTGECWTFAQLDAYSNAVANL 119
           ||||::      :    |      |:|:       |:   ||: :|: :       ||:
Db      32 PRIFRSKLPDIDIPNHLPLHTYCFEKLSSVSDKPCLIVGSTGKSYTYGETHLICRRVASG 91

Qy     120 FRQLGFAPGDVVAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRREPLAFCLGTSGAKAL 179
             :||   |||: | |:   |||  ::| :  |  :   |     : |   | :|||| :
Db      92 LYKLGIRKGDVIMILLQNSAEFVFSFMGASMIGAVSTTANPFYTSQELYKQLKSSGAKLI 151

Qy     180 IFGGEMVAAVAEVSGHLGKSLIKFCSGDLGPEGILPDTHLLDPLLKEASTAPLAQIPSKG 239
           |   : |  :     :||::|    : :  ||  ||       |: :  | |  :    |
Db     152 ITHSQYVDKLK----NLGENLTLITTDEPTPENCLP----FSTLITDDETNPFQETVDIG 203

Qy     240 MDDRLFYIYTSGTTGLPKAAIVVHSRYYRMAA-----FGHHAYRMQAADVLYDCLPLYHS 294
            ||     ::|||||||  :: |       |         : | ::: ||:   |||:|
Db     204 GDDAAALPFSSGTTGLPKGVVLTHKSLITSVAQQVDGDNPNLY-LKSNDVILCVLPLFH- 261

Qy     295 AGNIIGVGQCLI----YGLTVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVR 350
             |  :   |:      | ||:|  ||       |   ::  |:   :  :   | | |
Db     262 ---IYSLNSVLLNSLRSGATVLLMHKFEIGALLDLIQRHRVTIAALVPPLVIALAKNPTV 318

Qy     351 EAERRHRVRLAVGNG--LRPAIWEEFTERFGVRQIGEFYGATECNCSIANMDG------- 401
           :     ||  :     |   : :    |     :|: || ||    ::   |
Db     319 NSYDLSSVRFVLSGAAPLGKELQDSLRRRLPQAILGQGYGMTEAGPVLSMSLGFAKEPIP 378

Qy     402 -KVGSCGFNSRILPHVYPIRLVKVNEDTMELLRDAQG-LCIPCQAGEPGLLVGQINQQDP 459
            | ||||   |       :::| : |  :  |  :  | :||  |         || ::
Db     379 TKSGSCGTVVRNA----ELKVVHL-ETRLSLGYNQPGEICIRGQ---------QIMKE-- 422

Qy     460 LRRFDGYVSE-SATSKKIAHSVFSKGDSAYL-SGDVLVMDELGYMYFRDRSGDTFRWRGE 517
             |:::  ||| |       :  :| :||:  :||   ::  ||  :  :::|
Db     423 ------YLNDPEATSATI-------DEEGWLHTGDIGYVDEDDEIFIVDRLKEVIKFKGF 469

Qy     518 NVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVA 557
           |   |:| :|       | ||    ||  :  ||   ||
Db     470 QVPPAELESLLINHHSIADAAV----VPQNDEVAGEVPVA 505


RESULT 15
4CL1_ORYSA
ID   4CL1_ORYSA     STANDARD;      PRT;   563 AA.
AC   P17814;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   4-COUMARATE--COA LIGASE 1 (EC 6.2.1.12) (4CL 1) (4-COUMAROYL-COA
DE   SYNTHASE 1).
GN   4CL1 OR 4CL.
OS   Oryza sativa (Rice).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. JAPONICA;
RX   MEDLINE=91045096; PubMed=2235510;
RA   Zhao Y., Kung S.D., Dube S.K.;
RT   "Nucleotide sequence of rice 4-coumarate:CoA ligase gene, 4-CL.1.";
RL   Nucleic Acids Res. 18:6144-6144(1990).
CC   -!- CATALYTIC ACTIVITY: ATP + 4-COUMARATE + COA = AMP + DIPHOSPHATE +
CC       4-COUMAROYL-COA.
CC   -!- PATHWAY: BRANCH-POINT REACTIONS BETWEEN GENERAL PHENYLPROPANOID
CC       METABOLISM AND PATHWAYS LEADING TO VARIOUS SPECIFIC END PRODUCTS.
CC   -!- INDUCTION: TRANSCRIPTIONALLY ACTIVATED BY BOTH FUNGAL ELICITOR AND
CC       ULTRAVIOLET IRRADIATION.
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52623; CAA36850.1; -.
DR   PIR; JU0311; JU0311.
DR   HSSP; P08659; 1LCI.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Ligase; Phenylpropanoid metabolism; Multigene family.
SQ   SEQUENCE   563 AA;  60841 MW;  35B329401C6B44B9 CRC64;
```

```
  Query Match             8.3%;  Score 281;  DB 1;  Length 563;
  Best Local Similarity   26.0%;  Pred. No. 2.6e-14;
  Matches  149;  Conservative   87;  Mismatches  242;  Indels   94;  Gaps   29;

Qy      76 HTIPRIFQAVVQRQPE---RLALVDAGTGECWTFAQLDAYSNAVANLFRQ--LGFAPGDV 130
           :|:|  : :   :| ||   |  |:|  ||   |:| :|  |  :|   |:  ||   | |
Db      35 NTLP-LHRYCFERLPEVAARPCLIDGATGGVLTYADVDRLSRRLAAALRRAPLGLRRGGV 93

Qy     131 VAIFLEGRPEFVGLWLGLAKAGMEAALLNVNLRREP--LAFCLGTSGAKALIFGGEMVAA 188
           |   |   ||||  :   :: |  ||:   |    |  :   |  :||  :|   | :||
Db      94 VMSLLRNSPEFVLSFFAASRVG--AAVTTANPMSTPHEIESQLAAAGATVVI--TESMAA 149

Qy     189 ---VAEVSGHLGKSLIKFCSGDLGPEGILPDTHLLDPLLKEASTAPLA--QIPSK--GMD 241
              :  | |   ||     |   :| |   |  |.|: |   :|||  :   |    |
Db     150 DKLPSHSHGALTVVLI-----DERRDGCL---HFWDDLMSEDEASPLAGDEDDEKVFDPD 201

Qy     242 DRLFYIYTSGTTGLPKAAIVVHSRYYRMAA---------FGHHAYRMQAADVLYDCLPLY 292
           | :   |:||||||||  :: |       |         | |     | ||:   ||::
Db     202 DVVALPYSSGTTGLPKGVMLTHRSLSTSVAQQVDGENPNIGLH-----AGDVILCALPMF 256

Qy     293 H--SAGNIIGVGQCLIYGLTVVLRKKFSASRFWDDCIKYNCTVVQYIGEICRYLLKQPVR 350
           |  |   |:  |  |  |  :|: ::|  :   |   ::  |:   :  |   : |
Db     257 HIYSLNTIMMCG--LRVGAAIVVMRRFDLAAMMDLVERHRVTIAPLVPPIVVAVAKSEAA 314

Qy     351 EAERRHRVRLAVGNGLRPA---IWEEFTERFGVRQIGEFYGATECN------CSIANMDG 401
            |     ||: : :|  |    | : |  :     :|: || ||          : |
Db     315 AARDLSSVRMVL-SGAAPMGKDIEDAFMAKLPGAVLGQGYGMTEAGPVLSMCLAFAKEPF 373

Qy     402 KV--GSCGFNSRILPHVYPIRLVKVNEDTMELL-RDAQG-LCIPCQAGEPGLLVGQINQQ 457
           ||  |:||        |    |  :: || : | |: :| :||  |      :: | :|
Db     374 KVKSGACG------TVVRNAELKIIDPDTGKSLGRNLRGEICIRGQQ----IMKGYLNNP 423

Qy     458 DPLRRFDGYVSESATSKKI-AHSVFSKGDSAYLSGDVLVMDELGYMYFRDRSGDTFRWRG 516
           :              ||   | |      ||  |:  |    ||:   :  ||  :  ::||
Db     424 E------------ATKNTIDAEGWLHTGDIGYVDDD----DEI---FIVDRLKEIIKYRG 464

Qy     517 ENVSTTEVEGVLSRLLGQTDVAVYGVAVPGVEGKAGMAAVADPH-SLLDPNAIYQELQKV 575
             |:  |:| :|:       | || |:       |:  :| ||    | |  : : | : |
Db     465 FQVAPAELEALLNTHPSIADAAVVGLKF----GEIPVAFVAKTEGSELSEDDVKQFVAKE 520

Qy     576 LAPYARPIFLRLLPQVDTTGTFKIQKTRLQRE 607
           :  | :   :  : ::    : || :  |:::
Db     521 VIYYKKIREVFFVDKIPKAPSGKILRKELRKQ 552
```

```
Search completed: March  6, 2002, 13:42:25
Job time: 216 sec
```